A PRIMER ON SESQUILINEAR FORMS

BRIAN OSSERMAN

This is an alternative presentation of most of the material from §8.1, 8.2, 8.3, 8.4, 8.5 and 8.8 of Artin's book. Any terminology (such as sesquilinear form or complementary subspace) which is discussed here but not in Artin is optional for the course, and will not appear on homework or exams.

1. Sesquilinear forms

The dot product is an important tool for calculations in $\mathbb{R}^n$. For instance, we can use it to measure distance and angles. However, it doesn't come from just the vector space structure on $\mathbb{R}^n$ – to define it implicitly involves a choice of basis. In Math 67, you may have studied inner products on real vector spaces and learned that this is the general context in which the dot product arises. Now, we will make a more general study of the extra structure implicit in dot products and more generally inner products. This is the subject of bilinear forms. However, not all forms of interest are bilinear. When working with complex vector spaces, one often works with Hermitian forms, which toss in an extra complex conjugation. In order to handle both of these cases at once, we'll work in the context of sesquilinear forms.

For convenience, we'll assume throughout that our vector spaces are finite dimensional.

We first set up the background on field automorphisms.

Definition 1.1. Let F be a field. An **automorphism** of F is a bijection from F to itself which preserves the operations of addition and multiplication. An automorphism $\varphi : F \to F$ is an **involution** if $F \circ F$ is the identity map.

Example 1.2. Every field has at least one involution: the identity automorphism!

Example 1.3. Since for $z, z' \in C$, we have $\overline{z+z'} = \bar{z} + \bar{z}'$ and $\overline{zz'} = \bar{z}\bar{z}'$, complex conjugation is an automorphism. Since $\bar{\bar{z}} = z$, it is an involution.

Example 1.4. Consider the field $\mathbb{Q}(\sqrt{2}) = \{a + b\sqrt{2} : a, b \in \mathbb{Q}\}$. Then the map $a + b\sqrt{2} \mapsto a - b\sqrt{2}$ is an involution.

We will assume throughout that we are in the following

Situation 1.5. Let V be a vector space over a field F, and suppose we have an involution on F which we denote by $c \mapsto \bar{c}$ for all $c \in F$.

Definition 1.6. A **sesquilinear form** on a vector space V over a field F is a map

$$\langle , \rangle : V \times V \to F$$

which is linear on the right side, and almost linear on the left: that is,

$$\langle v_1, cw_1 \rangle = c \langle v_1, w_1 \rangle, \quad \text{and} \quad \langle v_1, w_1 + w_2 \rangle = \langle v_1, w_1 \rangle + \langle v_1, w_2 \rangle$$
$$\langle cv_1, w_1 \rangle = \bar{c} \langle v_1, w_1 \rangle, \quad \text{and} \quad \langle v_1 + v_2, w_1 \rangle = \langle v_1, w_1 \rangle + \langle v_2, w_1 \rangle$$

for all $c \in F$ and $v_1, v_2, w_1, w_2 \in V$.

A special case of sesquilinear forms that works for any field arises when the involution is the identity map.

Definition 1.7. If the chosen involution on F is the identity map, a sesquilinear form is called a **bilinear form**.

You should keep the bilinear case in mind as the main situation of interest throughout.

A simple induction implies that sesquilinear forms are compatible with arbitrary linear combinations, as follows:

Proposition 1.8. *If $\langle , \rangle$ is a sesquilinear form on V, then*

$$\left\langle \sum_{i=1}^{n} b_i v_i, \sum_{j=1}^{m} c_j w_j \right\rangle = \sum_{i,j} \bar{b}_i c_j \langle v_i, w_j \rangle$$

for any choices of $b_i, c_j \in F$ and $v_i, w_j \in V$.

We now examine how to express a sesquilinear form in terms of a matrix, given a choice of basis of V. A preliminary definition is the following.

Definition 1.9. Let $A = (a_{i,j})$ be a matrix with coefficients in F. The **adjoint** of A, written A^*, is obtained by transposing A and applying the chosen involution to its coefficients: that is,

$$A^* = (\bar{a}_{j,i}).$$

If a basis $(v_1, \ldots, v_n)$ of V is given, then every sesquilinear form on V can be expressed uniquely by a matrix, as follows: set $A = (a_{i,j})$ where

$$a_{i,j} = \langle v_i, v_j \rangle .$$

To go backwards from a matrix to a sesquilinear form (again, given the fixed choice of basis), we do the following:

Proposition 1.10. *Given a sesquilinear form $\langle , \rangle$ and a choice of basis $(v_1, \ldots, v_n)$ of V, define the matrix A as above. Then for any vectors $v = \sum_i b_i v_i$ and $w = \sum_i c_i v_i$ in V, we have*

$$\langle v, w \rangle = \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}^* A \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix}.$$

Proof. We expand out using the previous proposition.

$$\begin{aligned}
\langle v, w \rangle &= \left\langle \sum_i b_i v_i, \sum_j c_j v_j \right\rangle \\
&= \sum_{i,j} \bar{b}_i c_j \langle v_i, v_j \rangle \\
&= \sum_{i,j} \bar{b}_i c_j a_{i,j} \\
&= \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}^* \begin{bmatrix} a_{11} & \ldots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \ldots & a_{nn} \end{bmatrix} \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix}.
\end{aligned}$$

□

Example 1.11. If our involution is the identity map, this gives a correspondence between bilinear forms and matrices. In this case, the dot product associated to a given basis is simply the bilinear form corresponding to the identity matrix.

Since the matrix A depends on a choice of basis, it is natural to consider what happens if we change basis. This is described as follows.

Proposition 1.12. *Let $\langle , \rangle$ be a sesquilinear form on V. Let $B = (v_1, \ldots, v_n)$ and $B' = (v_1', \ldots, v_n')$ be two bases of V, and suppose that $\langle , \rangle$ is represented by the matrix A for the basis B, and by A' for the basis B'. If P is the change of basis matrix from B to B', so that $v_i' = \sum_{j=1}^{n} P_{j,i} v_j$ for $i = 1, \ldots, n$, then*

$$A' = P^* AP.$$

In particular, in the bilinear case we have $A = P^t AP$.

Proof. According to the definition, we need to verify that $\left\langle v_i', v_j' \right\rangle = (P^*AP)_{i,j}$. By Proposition 1.10, we have

$$\left\langle v_i', v_j' \right\rangle = \begin{bmatrix} P_{1,i} \\ \vdots \\ P_{n,i} \end{bmatrix}^* A \begin{bmatrix} P_{1,j} \\ \vdots \\ P_{n,j} \end{bmatrix},$$

but the righthand side is precisely $(P^*AP)_{i,j}$. ☐

Warning 1.13. Since, given a choice of basis of an n-dimensional vector space V, every bilinear form on V can be represented uniquely by an $n \times n$ matrix, and also every linear map from V to itself can be represented uniquely by an $n \times n$ matrix, one might think that the theory of bilinear forms is no different from the theory of linear maps from V to itself. However, they are not the same. We see this by considering what happens to the matrices in question if we change basis via an invertible matrix P. In the case of a linear map from V to itself given by a matrix A, the matrix changes to $P^{-1}AP$. However, by the previous proposition, if instead A is the matrix associated to a bilinear form, it changes to $P^t AP$.

A more abstract way of expressing the difference is that instead of a linear map $V \to V$, a bilinear form naturally gives a linear map $V \to V^*$, where V^* is the *dual space* of V defined as the collection of linear maps $V \to F$.

A natural condition to consider on sesquilinear forms is the following:

Definition 1.14. A sesquilinear form $\langle , \rangle$ is **symmetric** if $\langle v, w \rangle = \overline{\langle w, v \rangle}$ for all $v, w \in V$.

Note that unless the involution isn't the identity map, a "symmetric" sesquilinear form isn't quite symmetric, but it is as close as possible given the asymmetric nature of the definition of sesquilinear form. The special case of primary interest (other than bilinear forms) is the following:

Definition 1.15. If $F = \mathbb{C}$ and the involution is complex conjugation, then a symmetric sesquilinear form is called a **Hermitian form**.

Why did we introduce the complex conjugation rather than simply sticking with bilinearity? One reason is that with it, we see that for any $v \in V$, we have $\langle v, v \rangle = \overline{\langle v, v \rangle}$, so we conclude that $\langle v, v \rangle$ is fixed under complex conjugation, and is therefore a real number. This means we can impose further conditions by considering whether $\langle v, v \rangle$ is always positive, for instance. Another reason is that with this definition, if we multiply both sides by a complex number of length 1, such as i, we find that $\langle v, w \rangle$ doesn't change.

Example 1.16. On $\mathbb{C}^n$ we have the **standard Hermitian form** defined by

$$\langle (x_1, \ldots, x_n), (y_1, \ldots, y_n) \rangle = \bar{x}_1 y_1 + \cdots + \bar{x}_n y_n.$$

This has the property that $\langle v, v \rangle > 0$ for any nonzero v.

In fact, if we write $x_j = a_j + ib_j$ and $y_j = c_j + id_j$ (thus identifying $\mathbb{C}^n$ with $\mathbb{R}^{2n}$), we see that

$$\begin{aligned}\langle (x_1, \ldots, x_n), (y_1, \ldots, y_n) \rangle &= (a_1 - ib_1)(c_1 + id_1) + \cdots + (a_n - ib_n)(c_n + id_n) \\ &= (a_1c_1 + b_1d_1) + \cdots + (a_nc_n + b_nd_n) + i\left((a_1d_1 - b_1c_1) + \cdots + (a_nd_n - b_nc_n)\right).\end{aligned}$$

Notice that the real part of this is just the usual dot product on $\mathbb{R}^{2n}$. In particular, we get

$$\langle (x_1, \ldots, x_n), (x_1, \ldots, x_n) \rangle = (a_1^2 + b_1^2) + \cdots + (a_n^2 + b_n^2),$$

(the imaginary term cancels out). This is just the usual dot product on $\mathbb{R}^n$, which calculates the square of the length of a vector.

Here is a different example of a symmetric bilinear form.

Example 1.17. Using the standard basis on F^2, if we consider the bilinear form we obtain from the matrix $\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$, we see that

$$\left\langle \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}, \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} \right\rangle = b_1c_1 - b_2c_2.$$

This is still a symmetric form, but looks rather different from the dot product, since $\left\langle \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\rangle = -1$.

To study how symmetry of a sesquilinear form is reflected in its associated matrix, we define:

Definition 1.18. A matrix A is **self-adjoint** if $A^* = A$. In the special case that $F = \mathbb{C}$ and the involution is the identity map, a self-adjoint matrix is also called **Hermitian**.

Example 1.19. If the involution is the identity map, being self-adjoint is the same as being symmetric.

Just as with transpose, the adjoint satisfies the property that $(AB)^* = B^*A^*$ for any matrices A, B.

The following proposition describes how symmetry of a form carries over to its associated matrix.

Proposition 1.20. *Let $\langle , \rangle$ be a sesquilinear form described by a matrix A for a basis $(v_1, \ldots, v_n)$. Then the following are equivalent:*

(1) *$\langle , \rangle$ is symmetric;*
(2) *A is self-adjoint;*
(3) *For all i, j we have $\langle v_i, v_j \rangle = \overline{\langle v_j, v_i \rangle}$.*

Proof. If we write $A = (a_{i,j})$, then by definition $\langle v_i, v_j \rangle = a_{i,j}$ for all i, j, so it is immediate that (2) and (3) are equivalent.

Next, it is clear from the definition that (1) implies (3). Finally, if we assume (3), and we have vectors $v = \sum_i b_i v_i$ and $w = \sum_i c_i v_i$ in V, we have

$$\begin{aligned}\langle v, w \rangle &= \left\langle \sum_i b_i v_i, \sum_j c_j v_j \right\rangle \\ &= \sum_{i,j} \bar{b}_i c_j \left\langle v_i, \sum_j v_j \right\rangle.\end{aligned}$$

On the other hand,

$$\begin{aligned}
\overline{\langle w, v \rangle} &= \overline{\left\langle \sum_i c_i v_i, \sum_j b_j v_j \right\rangle} \\
&= \overline{\sum_{i,j} \bar{c}_i b_j \left\langle v_i, \sum_j v_j \right\rangle} \\
&= \sum_{i,j} c_i \bar{b}_j \overline{\left\langle v_i, \sum_j v_j \right\rangle},
\end{aligned}$$

and assuming (3), we see that this is equal to $\langle v, w \rangle$. Thus, we conclude that (3) implies (1), so all of the conditions are equivalent. ☐

We now discuss skew-symmetry. Although there is a notion of skew-symmetry for Hermitian forms, it has rather different behavior from the case of bilinear forms, so whenever we talk about skew-symmetry, we will restrict to the bilinear case.

Definition 1.21. A bilinear form is **skew-symmetric** if $\langle v, v \rangle = 0$ for all $v \in V$.

This terminology is justified by the following.

Proposition 1.22. *If $\langle , \rangle$ is a skew-symmetric bilinear form, then $\langle v, w \rangle = - \langle w, v \rangle$ for all $v, w \in V$. The converse is true if $0 \neq 2$ in F.*

Proof. If $\langle , \rangle$ is skew-symmetric, then for any v, w we have

$$0 = \langle v + w, v + w \rangle = \langle v, v \rangle + \langle v, w \rangle + \langle w, v \rangle + \langle w, w \rangle = \langle v, w \rangle + \langle w, v \rangle,$$

so $\langle v, w \rangle = - \langle w, v \rangle$.

Conversely, suppose that $\langle v, w \rangle = - \langle w, v \rangle$ for all $v, w \in V$. Then for any $v \in V$, we have $\langle v, v \rangle = - \langle v, v \rangle$, so $2 \langle v, v \rangle = 0$, and as long $2 \neq 0$ in F we can cancel the 2 to get $\langle v, v \rangle = 0$. ☐

Warning 1.23. We follow Artin's terminology for skew-symmetric forms (see §8.8). However, some other sources (such as, for instance, Wikipedia) define a form to be skew-symmetric if $\langle v, w \rangle = - \langle w, v \rangle$ for all $v, w \in V$, and call it *alternating* if $\langle v, v \rangle = 0$ for all $v \in V$. Of course, according to the proposition these are the same as long as $2 \neq 0$ in F.

Example 1.24. Still using the standard basis on F^2, if we consider the bilinear form we obtain from the matrix $\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$, we see that

$$\left\langle \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}, \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} \right\rangle = b_1 c_2 - b_2 c_1.$$

This is now a skew-symmetric form.

In general, we have the following proposition.

Proposition 1.25. *Let $\langle , \rangle$ be a bilinear form described by a matrix A for a basis $(v_1, \dots, v_n)$. Then the following are equivalent:*

(1) *$\langle , \rangle$ is skew-symmetric;*
(2) *$A^t = -A$ and the diagonal entries of A are equal to 0;*
(3) *For all i, j we have* $\langle v_i, v_j \rangle = \begin{cases} - \langle v_j, v_i \rangle : & i \neq j \\ 0 : & i = j. \end{cases}$

Notice that this proposition gives a different proof of Proposition 1.22.

Proof. If we write $A = (a_{i,j})$, we have $\langle v_i, v_j \rangle = a_{i,j}$ by definition, so it is clear that (2) and (3) are equivalent.

We have just proved that if $\langle , \rangle$ is skew-symmetric, then $\langle v_i, v_j \rangle = -\langle v_j, v_i \rangle$, and $\langle v_i, v_i \rangle = 0$ follows directly from the definition, so we conclude that (1) implies (3).

Conversely, suppose that (3) is satisfied. Then if $v = \sum_i b_i v_i$, we have

$$\begin{aligned}\langle v, v \rangle &= \left\langle \sum_i b_i v_i, \sum_j b_j v_j \right\rangle \\ &= \sum_{i,j} b_i b_j \langle v_i, v_j \rangle \\ &= \sum_{i<j} b_i b_j (\langle v_i, v_j \rangle + \langle v_j, v_i \rangle) + \sum_i b_i b_i \langle v_i, v_i \rangle \\ &= 0.\end{aligned}$$

Thus, we see that (3) implies (1), so all the conditions are equivalent. □

2. Hermitian forms and reality

Now we take a brief interlude to explore what makes Hermitian forms special. Let's recall a few more basic linear algebra definitions:

Definition 2.1. Let A be an $n \times n$ matrix with coefficients in a field F. A nonzero vector $v \in F^n$ is an **eigenvector** of A if $Av = \lambda v$ for some $\lambda \in F$. We call λ an **eigenvalue** of A.

The **trace** of A is the sum of the diagonal entries.

The **characteristic polynomial** of A is the polynomial in λ given by $\det(\lambda I - A)$.

Proposition 2.2. *The eigenvalues of A are the roots of its characteristic polynomial. The constant term of the characteristic polynomial is $(-1)^n \det A$, and the second coefficient is negative the trace of A.*

In particular, the trace and the determinant of A can be expressed as a sum and as a product of eigenvalues, respectively.

We now return to the special case of complex vector spaces, and prove an important property of Hermitian matrices.

Theorem 2.3. *If A is a Hermitian matrix, then the eigenvalues, determinant, and trace of A are all real numbers.*

Proof. Since the determinant and trace are products and sums of eigenvalues, it is enough to prove the eigenvalues are real. Suppose $v = \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}$ is an eigenvector of A. Then $Av = \lambda v$ for some λ, and we want to prove that λ is real. We know that $\langle v, v \rangle = v^* A v = v^* \lambda v = \lambda v^* v$. On the other hand, we know that $\langle v, v \rangle$ is a real number, and $v^* v$ is a nonzero real number (since it is just the standard Hermitian form applied to v), so we conclude that λ is real, as desired. □

Since real symmetric matrices are a special case of Hermitian matrices, we conclude the following:

Corollary 2.4. *If A is a real symmetric matrix, all the eigenvalues of A are real.*

What do we mean by this? We mean that even allowing for going from real to complex numbers, we will not obtain any non-real eigenvalues of a real symmetric matrix. Equivalently, all complex roots of the characteristic polynomial are in fact real. We will prove a stronger version of this statement soon.

Note that even though the statement of the corollary is just in terms of real numbers, it is natural to work with complex vectors (and therefore Hermitian forms) in proving the statement, since one has to consider the possibility of complex eigenvectors and eigenvalues in order to prove that in the end, everything is real.

3. Orthogonality (§8.4)

In this section, we have the following basis situation:

Situation 3.1. We have a sesquilinear $\langle , \rangle$ with the property that $\langle v, w\rangle = 0$ if and only if $\langle w, v\rangle = 0$.

This condition is satisfied if the form is symmetric (even if the general sesquilinear sense), or if it is a skew-symmetric bilinear form. Later, we will specialize to the symmetric case.

Definition 3.2. Two vectors v, w are **orthogonal** (written $v \perp w$) if $\langle v, w\rangle = 0$.

Thus, by hypothesis we have $\langle v, w\rangle = 0$ if and only if $\langle w, v\rangle = 0$.

In this generality, we may well have $v \perp v$ even if $v \neq 0$ (indeed, this will always be the case for skew-symmetric forms, but can occur also in the symmetric or Hermitian cases). Thus, it is better not to try to place too much geometric significance on the notion of orthogonality, even though it is very useful.

Definition 3.3. If $W \subseteq V$ is a subspace, the **orthogonal space** $W^\perp$ to W is defined by

$$W^\perp = \{v \in V : \forall w \in W, v \perp w\};$$

this is a subspace of V.

Definition 3.4. A basis $(v_1, \dots, v_n)$ of V is **orthogonal** if $v_i \perp v_j$ for all $i \neq j$.

Definition 3.5. A **null vector** in V is a vector which is orthogonal to every $v \in V$. The **nullspace** is a set (which is a subspace) of all null vectors.

We see that the nullspace can also be described as $V^\perp$.

Definition 3.6. The form $\langle , \rangle$ is **nondegenerate** if its nullspace is $\{0\}$. If the form is not nondegenerate, it is degenerate.

Thus, the form is nondegenerate if and only if for every nonzero $v \in V$, there is some $v' \in V$ with $\langle v, v'\rangle \neq 0$.

Definition 3.7. Given a subspace $W \subseteq V$, the form $\langle , \rangle$ is **nondegenerate on** W if $W^\perp \cap W = \{0\}$.

Thus, $\langle , \rangle$ is nondegenerate on W if, for every nonzero $w \in W$, there is some $w' \in W$ such that $\langle w, w'\rangle \neq 0$. We may reexpress this as follows: we can define the **restriction** of $\langle , \rangle$ to W to be the form on W obtained by the inclusion of W into V, forgetting what happens for vectors not in W. Then $\langle , \rangle$ is nondegenerate on W if and only if the restriction to W is a nondegenerate form.

The following is immediate from the definitions:

Proposition 3.8. *The matrix of the form with respect to a basis B is diagonal if and only if B is orthogonal, and in this case the form is nondegenerate if and only if none of the diagonal entries are equal to* 0.

Remark 3.9. Note that this means that a nonzero skew-symmetric form can never have an orthogonal basis, since the matrix for a skew-symmetric form always has all 0s on the diagonal.

The following proposition gives us a method for testing when two vectors are equal.

Proposition 3.10. *Suppose that $\langle , \rangle$ is nondegenerate, and $v, v' \in V$. If $\langle v, w\rangle = \langle v', w\rangle$ for all $w \in V$, then $v = v'$.*

Proof. If $\langle v, w\rangle = \langle v', w\rangle$, then $\langle v - v', w\rangle = 0$, so $(v - v') \perp w$. If this is true for all $w \in V$, we conclude that $v - v' \in V^\perp$, so by nondegeneracy of $\langle , \rangle$, we must have $v - v' = 0$. $\square$

Next we relate null vectors and nondegeneracy to the matrix describing the form.

Proposition 3.11. *Suppose that we have a basis $(v_1, \ldots, v_n)$ of V, and A is the matrix for $\langle , \rangle$ in terms of this basis. Then:*

(1) *A vector $v = \sum_i b_i v_i$ is a null vector if and only if*

$$A \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix} = 0.$$

(2) *The form $\langle , \rangle$ is nondegenerate if and only if A is invertible.*

Proof. (1) Since for $w = \sum_i c_i v_i$, we have

$$\langle w, v\rangle = \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix}^* A \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix},$$

it is clear that if $A \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix} = 0$, then v is a null vector. Conversely, if $w = w_i$, then we see that $\langle w, v\rangle$ is equal to the ith coordinate of $A \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}$, so if v is a null vector, we conclude that each coordinate of $A \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}$ is equal to 0, and therefore that $A \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix} = 0$.

(2) By definition, $\langle , \rangle$ is nondegenerate if and only if there is no nonzero null vector, and by part (1) this is equivalent to $AY = 0$ not having any nonzero solutions (Y a $n \times 1$ column vector). But this in turn is equivalent to A being invertible, since it is a square matrix. $\square$

Since we have not yet specialized to the symmetric/Hermitian case, the following theorem includes both Theorem 8.4.5 and Theorem 8.8.6 of Artin. I have attempted to give a more conceptual proof than Artin does.

Theorem 3.12. *Let $W \subseteq V$ be a subspace.*

(1) *$\langle , \rangle$ is nondegenerate on W if and only if V is the direct sum $W \oplus W^\perp$.*
(2) *If $\langle , \rangle$ is nondegenerate on V and on W, then it is nondegenerate on $W^\perp$.*

If $W_1, W_2 \subseteq V$ are subspaces, what does it mean to say V is $W_1 \oplus W_2$? The more abstract way to say it is that we always have the vector space $W_1 \oplus W_2$, and this always has a natural map to V, coming from the inclusions of W_1 and W_2 into V. Namely, if $(w_1, w_2) \in W_1 \oplus W_2)$, then it maps to $w_1 + w_2$ in V. We say that V is $W_1 \oplus W_2$ if this natural map is an isomorphism. More concretely, this is the same thing as saying that every vector in V can be written uniquely as $w_1 + w_2$ for

$w_1 \in W_1$, $w_2 \in W_2$. This breaks down into two statements: first, that $W_1 \cap W_2 = \{0\}$, and second, that $W_1 + W_2 = V$.

We will prove a couple of preliminary facts before giving the proof of the theorem.

Proposition 3.13. *If $W \subseteq V$ is a subspace, and $(w_1, \dots, w_m)$ is a basis of W, then $v \in V$ is in $W^\perp$ if and only if $v \perp w_i$ for $i = 1, \dots, m$.*

Proof. Certainly if $v \in W^\perp$, then $v \perp w_i$ for all i. Conversely, if $v \perp w_i$ for all i, then for any $w \in W$, write $w = \sum_i c_i w_i$, and we have

$$\langle v, w \rangle = \sum_i c_i \langle v, w_i \rangle = 0.$$

□

Lemma 3.14. *If $W \subseteq V$ is a subspace, then*

$$\dim W^\perp \geqslant \dim V - \dim W.$$

Proof. Suppose $(w_1, \dots, w_m)$ is a basis of W. We use $\langle , \rangle$ to construct a linear map $\varphi : V \to F^m$ by sending v to $(\langle w_1, v \rangle, \dots, \langle w_m, v \rangle)$. Since $\langle , \rangle$ is linear on the righthand side, this is indeed a linear map. Moreover, we see that the kernel of the map is precisely the set of vectors $v \in V$ such that $v \perp w_i$ for all i, which by Proposition 3.13 is exactly $W^\perp$. Now, the image of this map is contained in F^m, so has dimension at most m. Since $\dim \ker \varphi + \dim \operatorname{im} \varphi = \dim V$, we conclude $\dim W^\perp + \dim \operatorname{im} \varphi = \dim V$, so

$$\dim W^\perp \geqslant \dim V - m = \dim V - \dim W,$$

as desired. □

Proof of Theorem 3.12. For (1), first recall that $\langle , \rangle$ is nondegenerate on W if and only if $W \cap W^\perp = \{0\}$. Thus, if $V = W \oplus W^\perp$, then $\langle , \rangle$ is nondegenerate on W, and to prove the converse, we need to check that if $\langle , \rangle$ is nondegenerate, then the map $W \oplus W^\perp \to V$ is an isomorphism. But the kernel of this map is vectors of the form $(w, -w)$, where $w \in W \cap W^\perp$, so if $\langle , \rangle$, we have that the map is injective. Thus, we have that the dimension of the image is equal to $\dim(W \oplus W^\perp) = \dim W + \dim W^\perp$. But by Lemma 3.14, this is at least $\dim W + \dim V - \dim W = \dim V$. On the other hand, the image is contained in V, so it can have dimension at most $\dim V$, and we conclude that the image dimension is exactly $\dim V$, and hence that the image is all of V, which is what we wanted to show.

For (2), given any nonzero $w \in W^\perp$, we wish to show that there is some $w' \in W^\perp$ such that $\langle w, w' \rangle \neq 0$. Because we have assumed nondegeneracy on V, there is some $v \in V$ such that $\langle w, v \rangle \neq 0$. By (1) (using nondegeneracy on W), we can write $v = v' + w'$, where $v' \in W$, and $w' \in W^\perp$. But then since $v' \in W$ and $w \in W^\perp$, we have

$$0 \neq \langle w, v \rangle = \langle w, v' + w' \rangle = \langle w, v' \rangle + \langle w, w' \rangle = 0 + \langle w, w' \rangle = \langle w, w' \rangle,$$

as desired. □

Here is a different point of view on the first part of the theorem: it says that if a form $\langle , \rangle$ is nondegenerate on W, then we always have a natural projection from V to W. Here are the relevant definitions:

Definition 3.15. If $W \subseteq V$ is a subspace, a **projection** $\pi : V \to W$ is a linear map such that $\pi(w) = w$ for all $w \in W$.

A **complementary subspace** for W is a subspace $W' \subseteq V$ such that $V = W \oplus W'$.

Thus, the theorem says if $\langle , \rangle$ is nondegenerate on W, then $W^\perp$ is a complementary subspace for W. This is related to projections as follows:

Proposition 3.16. *Let $W \subseteq V$ be a subspace. Given a projection $\pi : V \to W$, the kernel of π is a complementary subspace. On the other hand, given a complementary subspace W', the map $V \to W$ obtained by writing $V = W \oplus W'$ and mapping $v = (w, w')$ to w is a projection. These two procedures are inverse to one another, and give a bijection between projections $V \to W$ and complementary subspaces $W' \subseteq V$ for W.*

We leave the proof of this proposition to the reader.

In general, for a given subspace there is no natural choice of projection to it, since there are many complementary subspaces. However, with additional structure such as a nondegenerate form, we can obtain a natural projection. Indeed, from Theorem 3.12 together with Proposition 3.16, we conclude:

Corollary 3.17. *If $\langle , \rangle$ is nondegenerate on W, then it induces a projection $V \to W$, with kernel equal to $W^\perp$.*

Such a projection is called the **orthogonal projection**. We will describe it in more detail in the symmetric case. In the case of the dot product on $\mathbb{R}^n$, orthogonal projection is the map which sends a vector to the closest point which lies in W. In full generality however, there is no such geometric intuition for it.

4. Symmetric forms

We continue with our discussion of orthogonality, but now we specialize to the case that $\langle , \rangle$ is symmetric. We further assume that $2 \neq 0$ in F. Our first order of business is to prove existence of orthogonal bases.

Note that the following lemma is *always* false if we consider instead a skew-symmetric form.

Lemma 4.1. *If $\langle , \rangle$ is not the zero form, there is some $v \in V$ such that $\langle v, v \rangle \neq 0$.*

Proof. Since $\langle , \rangle$ is not the zero form, there is some pair of vectors $w, w' \in V$ such that $\langle w, w' \rangle \neq 0$. We may further assume that $\langle w, w' \rangle = \langle w', w \rangle$ – in the case that $\langle , \rangle$ is bilinear, this is automatic, while in the general case, if $\langle w, w' \rangle = c$, if we replace w' by $\frac{1}{c} w'$, we will have $\langle w, w' \rangle = 1$. Since our involution preserves multiplication, we must have $\bar{1} = 1$, so we get $\langle w', w \rangle = \langle w, w' \rangle$, as desired.

Then we have
$$\langle w + w', w + w' \rangle = \langle w, w \rangle + 2 \langle w, w' \rangle + \langle w', w' \rangle ,$$
and since $2 \neq 0$ we have that $2 \langle w, w' \rangle \neq 0$, so at lease one other term in the equation must be nonzero. Thus we see that we can get the desired v as (at least one of) w, w' or $w + w'$. □

Theorem 4.2. *$\langle , \rangle$ has an orthogonal basis.*

Proof. We prove the statement by induction on $\dim V$. For the base case $V = 0$, we may use the empty basis. Suppose then that $\dim V = n$, and we know the theorem for any vector spaces of dimension less than n. If $\langle , \rangle$ is the zero form, then any basis is orthogonal. Otherwise, by Lemma 4.1, there is some $v \in V$ with $\langle v, v \rangle \neq 0$. Then let W be the subspace of V spanned by v. Since this is 1-dimensional and $\langle v, v \rangle \neq 0$, we see that $\langle , \rangle$ is nondegenerate on W. By Theorem 3.12, we have $V = W \oplus W^\perp$. Now, $W^\perp$ has dimension $n - 1$, so by induction it has an orthogonal basis $(v_1, \ldots, v_{n-1})$. But then if we set $v_n = v$, we see that we get an orthogonal basis of V. □

Remark 4.3. Although we pointed out that as a consequence of Proposition 3.8, a nonzero skew-symmetric form can never have an orthogonal basis, there is a sort of substitute which allows us to put the matrix for the form into a standard, close-to-diagonal form. This is off topic for us, but see Theorem 8.8.7 of Artin if you're curious.

We now examine how orthogonal bases give us explicit projection formulas.

Theorem 4.4. *Let $W \subseteq V$ be a subspace on which $\langle , \rangle$ is nondegenerate, and suppose $(w_1, \ldots, w_k)$ is an orthogonal basis of W. Then the orthogonal projection $\pi : V \to W$ is given by $\pi(v) = \sum_i c_i w_i$, where*

$$c_i = \frac{\langle w_i, v \rangle}{\langle w_i, w_i \rangle}.$$

Proof. First note that $\langle w_i, w_i \rangle \neq 0$ because the form is nondegenerate. In order to show that the given formula describes the orthogonal projection, it suffices to check that it is a projection, and to check that the kernel is equal to $W^\perp$. The map is question is visibly linear, so to check it is a projection we just check that it sends $w = \sum_i b_i w_i$ to itself. But then

$$\begin{aligned} c_i &= \frac{\sum_j b_j \langle w_i, w_j \rangle}{\langle w_i, w_i \rangle} \\ &= \frac{b_i \langle w_i, w_i \rangle}{\langle w_i, w_i \rangle} \\ &= b_i, \end{aligned}$$

since the w_i form an orthogonal basis.

Finally, by Proposition 3.13, we know that $v \in V$ is in $W^\perp$ if and only if $\langle w_i, v \rangle = 0$ for all i, so we see that the kernel of the map is exactly $W^\perp$, as desired. □

Projection from V onto itself is not very interesting – by definition, it has to be the identity map! However, the formula of Theorem 4.4 is still interesting in the case $W = V$.

Corollary 4.5. *Suppose that $\langle , \rangle$ is nondegenerate on V, and $(v_1, \ldots, v_n)$ is an orthogonal basis. Then for any $v \in V$, we have*

$$v = \sum_i \frac{\langle v_i, v \rangle}{\langle v_i, v_i \rangle} v_i.$$

That is, for any orthogonal basis, there is a simple method of finding how to express any vector in terms of the basis, just using the given form. (Compare to the usual case, where to figure out how to express a vector in terms of a basis involves inverting the change-of-basis matrix)

Example 4.6. Say we want to find the formula for orthogonal projection to the line $(t, 2t, 3t)$ in $\mathbb{R}^3$ under dot product. That is, given (x, y, z), give a formula for the value of t such that $(t, 2t, 3t)$ is cloest to (x, y, z). We have implicitly chosen $(1, 2, 3)$ as the basis for the line, so according to Theorem 4.4, we find that the value of t we want is

$$\frac{\langle (1,2,3), (x,y,z) \rangle}{\langle (1,2,3), (1,2,3) \rangle} = \frac{x + 2y + 3z}{1 + 4 + 9} = \frac{x + 2y + 3z}{14}.$$

Definition 4.7. An **orthonormal basis** is an orthogonal basis $(v_1, \ldots, v_n)$ such that $\langle v_i, v_i \rangle = 1$ for all i.

Thus, a given bilinear form has an orthonormal basis if and only if it can be thought of as being the dot product with respect to that basis (and for Hermitian forms, the same is true with the standard Hermitian form in place of the dot product).

Note that given an orthonormal basis, the formulas of Theorem 4.4 and Corollary 4.5 simplify further, because the denominators are all equal to 1.

We now specialize even further, to the case that we either have a real symmetric bilinear form, or a Hermitian form. In this situation, we can normalize our basis further. We first define:

Definition 4.8. A real symmetric bilinear form or Hermitian form is **positive definite** if $\langle v, v \rangle > 0$ for all nonzero $v \in V$.

We then have:

Corollary 4.9. *If $\langle , \rangle$ is a real symmetric bilinear form or a Hermitian form on V, then there exists an orthogonal basis $(v_1, \dots, v_n)$ such that $\langle v_i, v_i \rangle$ is either 1, -1 or 0.*

We have $\langle , \rangle$ positive definite if and only if it has an orthonormal basis.

Proof. If we scale v_i by c, then $\langle v_i, v_i \rangle$ scales by c^2 in the real case, and by $\bar{c}c$ in the Hermitian case. Thus, we can scale by an arbitrary positive real number. Since $\langle v_i, v_i \rangle$ starts off as a real number, we get what we want.

For the next assertion, if $(v_1, \dots, v_n)$ is an orthonormal basis, and $v = \sum_i c_i v_i$, then

$$\langle v, v \rangle = \begin{cases} \sum_i \ c_i^2 : & \text{real case} \\ \sum_i \bar{c}_i c_i : & \text{Hermitian case.} \end{cases}$$

In either case, it is visibly positive definite.

Conversely, if $\langle , \rangle$ is positive definitive, and $(v_1, \dots, v_n)$ is a basis as in Corollary 4.9, then we see that we must have $\langle v_i, v_i \rangle = 1$ for all i, so $(v_1, \dots, v_n)$ is an orthornomal basis. $\square$

Thus, being positive definite is equivalent to being the dot product (respectively, standard Hermitian form) with respect to some basis.

Finally, we also see that for positive definite forms, we don't need to worry about degeneracy:

Corollary 4.10. *If $\langle , \rangle$ is a positive definite real symmetric bilinear or Hermitian form on V, then for every subspace $W \subseteq V$, we have $\langle , \rangle$ nondegenerate on W, and in particular, $V = W \oplus W^\perp$.*

Proof. We see directly that the only null vector in W is 0: if $w \in W$ is nonzero, then by the definition of positive definite, we have $\langle w, w \rangle \neq 0$, so w is not a null vector. Thus, $\langle , \rangle$ is nondegenerate on W, so $V = W \oplus W^\perp$ by Theorem 3.12. $\square$